



Interpretability suprema in Peano Arithmetic

Paula Henk¹ · Albert Visser²

Received: 28 January 2015 / Accepted: 12 May 2017 / Published online: 26 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract This paper develops the philosophy and technology needed for adding a supremum operator to the interpretability logic ILM of Peano Arithmetic (PA). It is well-known that any theories extending PA have a supremum in the interpretability ordering. While provable in PA, this fact is not reflected in the theorems of the modal system ILM, due to limited expressive power. Our goal is to enrich the language of ILM by adding to it a new modality for the interpretability supremum. We explore different options for specifying the exact meaning of the new modality. Our final proposal involves a unary operator, the dual of which can be seen as a (nonstandard) provability predicate satisfying the axioms of the provability logic GL.

Keywords Provability logic · Interpretability · Peano Arithmetic

Mathematics Subject Classification 03F45 · 03F25 · 03F30

✉ Paula Henk
p.henk@uva.nl
Albert Visser
a.visser@uu.nl

¹ Institute for Logic, Language, and Information, University of Amsterdam, Science Park 107,
1098 XG Amsterdam, The Netherlands

² Philosophy, Faculty of Humanities, Utrecht University, Janskerkhof 13, 3512 BL Utrecht,
The Netherlands

1 Introduction

The aim of this paper is to bring closer together the two main approaches to the study of interpretability. As we will explain below, one of them currently lacks expressive power for properly talking about the other.

Given first-order theories T and S , possibly in different languages, we say that T interprets S , and write $T \triangleright S$, if there is some structure-preserving translation from the language of S to the language of T , such that the translations of all theorems of S are provable in T .

Interpretability can be seen as *generalized provability*: T is required to prove everything that S proves modulo some well-behaved translation. As such, it allows us to *compare* theories talking about different objects, such as Peano Arithmetic (PA) and Zermelo-Fraenkel set theory (ZF). In fact, the notion of interpretability is a natural candidate for giving a precise meaning to the intuitive idea of one theory being stronger than another one. Seen from a semantic perspective, an interpretation of S in T gives rise to a uniform way of constructing a model of S inside a given model of T . Interpretations therefore give rise to relative consistency proofs.

The study of interpretability may roughly be divided into two traditions, briefly outlined below. For a more comprehensive overview, the reader is referred to [32].

The lattice of degrees Seeing interpretability as a measuring stick for theories naturally leads one to study the space of all theories ordered by this relation. A *degree* is a collection of all theories that are equally strong as a given theory, i.e. that all mutually interpret each other. We write $[S]$ for the degree of S .

Among the first results concerning interpretability degrees is a strengthening of Gödel's Second Incompleteness Theorem by Feferman. Denote by Con_{PA} the usual consistency statement of PA. In [4, Theorem 6.5], it is shown that not only is Con_{PA} unprovable, it is also "uninterpretable": $[\text{PA}] \not\triangleright [\text{PA} + \text{Con}_{\text{PA}}]$. Jeroslow [12, Theorem 3.1, 3.2] showed that the degrees intermediate between $[\text{PA}]$ and $[\text{PA} + \text{Con}_{\text{PA}}]$ form a dense partial order. It follows from his work that the interpretability ordering is dense (see [28, p.798]). Montague [20, Theorem 1] proved the existence of an infinite set of finitely axiomatized subtheories of PA, all of which are mutually incomparable with respect to the interpretability ordering.

A systematic study of interpretability degrees was undertaken independently by Švejdar and Lindström [7, p.402]. Švejdar studied the structure $(\mathcal{V}, \triangleright)$ of the degrees of finite extensions of PA. He proved, among other things, that this structure is a distributive lattice [28, Theorems 4.4, 4.7]. Lindström was concerned with the structure of the degrees of *all* r.e. extensions of PA [15, 16]. However he also showed that this structure is in fact isomorphic to $(\mathcal{V}, \triangleright)$ [16, p.348, Theorem 3].

Our interest in provability logic makes it natural to focus on $(\mathcal{V}, \triangleright)$. Note first that $[\text{PA} + \top]$ is the minimum and $[\text{PA} + \perp]$ the maximum element of this structure.¹ It is easy to see that the infimum of $[\text{PA} + A]$ and $[\text{PA} + B]$ in $(\mathcal{V}, \triangleright)$ is $[\text{PA} + (A \vee B)]$. However the supremum of $[\text{PA} + A]$ and $[\text{PA} + B]$ is in general *not* $[\text{PA} + (A \wedge B)]$.

¹ Someone used to Boolean algebras has to think of the lattice $(\mathcal{V}, \triangleright)$ as being *upside down*: the weakest theory \top is the bottom element, and the strongest theory \perp is the top element of $(\mathcal{V}, \triangleright)$.

Švejdar shows that the supremum of $[PA + A]$ and $[PA + B]$ may taken to be $[PA + \vartheta]$, where ϑ is obtained by the Diagonal Lemma as a sentence such that:²

$$PA \vdash \vartheta \leftrightarrow \forall x (\text{Con}_x(\vartheta) \rightarrow \text{Con}_x(A) \wedge \text{Con}_x(B)). \tag{1}$$

Interpretability logic Interpretability, like provability, is a syntactical notion, and can therefore be formalized in the language of arithmetic. We can thus ask: which statements concerning provability and interpretability (between its finite extensions) are provable in PA?

It turns out that using *modal logic* allows one to answer this question in a neat way. The system GL of propositional provability logic contains a unary modality \Box , interpreted as formalized provability in PA. It follows from the work of Hilbert and Bernays [11], Löb [18], and Solovay [27], that the theorems of GL are exactly the propositional schemata involving formalized provability that are provable in PA.

The system ILM of interpretability logic is obtained by adding to GL a binary modality \triangleright for interpretability. It was proven independently by Berarducci [1] and Shavrukov [24] that the theorems of ILM are exactly the propositional schemata involving formalized provability and interpretability that are provable in PA.

Plan of this paper Given the two traditions in the study of interpretability, it is natural to ask how they relate to each other. The starting point for this paper is the following question: does the modal logic ILM “know” that $(\mathcal{V}, \triangleright)$ is a lattice?

The fact that the infimum in $(\mathcal{V}, \triangleright)$ is given by disjunction is indeed reflected in the axioms of ILM. On the other hand, Švejdar’s construction of a supremum employs a language much more complex than that available in propositional modal logic. Indeed, the supremum turns out not to be definable in ILM.

We want to boost the expressive power of ILM by adding to it a *new modality* whose intended interpretation is a supremum operator in $(\mathcal{V}, \triangleright)$. As we will see, Švejdar’s construction is only one out of many ways of obtaining the supremum of given elements in this structure. Since each of these can, in principle, be used to specify the meaning of the new modality, we are faced with a design choice.

Section 3 discusses various considerations to be taken into account in the midst of this embarrassment of riches. Švejdar’s construction, along with its dual discovered by Visser, are studied in Sect. 4. Section 5 introduces our favourite way of interpreting the new modality: a combination of conjunction with a certain unary operator that we call a *supremum adapter*. The idea and first examples of supremum adapters are due to Shavrukov. Supremum adapters may be seen as certain nonstandard provability predicates. This perspective is explored in Sect. 6, where we study the bimodal logic of such a nonstandard notion of provability, together with ordinary provability.

² $\text{Con}_x(C)$ denotes the consistency statement for $I\Sigma_x + C$, where $I\Sigma_x$ is as usual PA with induction restricted to Σ_x -formulas. In Švejdar’s original construction, the sentence $\text{Con}_x(C)$ is taken to be the consistency statement for C together with the first x axioms of PA. For our purposes, it is more convenient to use the stratification sequence based on $I\Sigma_x$ instead.

2 Preliminaries

This section introduces the basic notions used in the paper. The reader is advised to go through it lightly in order to return when some fact or definition is used.

2.1 Arithmetical theories

We consider first-order theories in the language \mathcal{L} of arithmetic containing $0, \mathbf{S}$ (successor), $+, \times$, and \leq . As usual, a formula is said to be Δ_0 (equivalently: Σ_0 or Π_0) if all its quantifiers are bounded. A formula is Σ_{n+1} (Π_{n+1}) if it is of the form $\exists x_0 \dots x_n A$, with A a Π_n (Σ_n)-formula. For each natural number n , we define a term \bar{n} of \mathcal{L} by letting $\bar{0} = 0$, and $\overline{n+1} = \mathbf{S}\bar{n}$. Given that, we will often write n instead of \bar{n} .

The basic facts concerning $0, \mathbf{S}, +, \times$, and \leq are given by the axioms of the theory \mathbf{Q} of Robinson Arithmetic [7, Definition I.1.1]. The theory \mathbf{Q} is Σ_1 -complete, meaning that it proves every true Σ_1 -sentence. The theory \mathbf{PA} of Peano Arithmetic results from adding to \mathbf{Q} the induction schema for all \mathcal{L} -formulas.

As usual, $\mathbf{I}\Sigma_n$ is the fragment of \mathbf{PA} obtained by restricting the induction schema to Σ_n -formulas. Note that $\mathbf{I}\Sigma_n \subseteq \mathbf{I}\Sigma_{n+1}$ for all n . We recall that for $n > 0$, $\mathbf{I}\Sigma_n$ is finitely axiomatizable [7, Theorem I.2.52]. The theory $\mathbf{I}\Delta_0 + \mathbf{exp}$ results from adding to $\mathbf{I}\Delta_0$ an axiom stating the totality of exponentiation. We note that $\mathbf{I}\Delta_0 + \mathbf{exp}$ is finitely axiomatizable [7, Theorem V.5.6] and contained in $\mathbf{I}\Sigma_1$ [7, Theorem I.1.50].

Metamathematics It is well-known that arithmetization of syntax can be carried out in $\mathbf{I}\Delta_0 + \mathbf{exp}$. We assume as given some standard gödelnumbering of \mathcal{L} -formulas, and write $\ulcorner A \urcorner$ for the gödelnumber of A . We shall often identify a formula with its gödelnumber, writing for example $B(A)$ instead of $B(\ulcorner A \urcorner)$.

Let S be a recursively enumerable (r.e.) extension of $\mathbf{I}\Delta_0 + \mathbf{exp}$. Since any r.e. relation can be represented in a natural way in $\mathbf{I}\Delta_0 + \mathbf{exp}$ by a Σ_1 -formula [3], there is a Σ_1 -formula σ representing the axioms of S in $\mathbf{I}\Delta_0 + \mathbf{exp}$, i.e. we have

$$A \text{ is an axiom of } S \quad \text{iff} \quad \mathbf{I}\Delta_0 + \mathbf{exp} \vdash \sigma(A).$$

Using σ , one can define in a natural way (see [4, Definition 4.1]) a Σ_1 -formula Pr_σ representing (inside $\mathbf{I}\Delta_0 + \mathbf{exp}$) provability in S . In this paper, we will write Pr_S instead of Pr_σ , having in mind some standard Σ_1 -representation σ of the axioms of S in $\mathbf{I}\Delta_0 + \mathbf{exp}$. To further simplify notation, we shall write \Box_S for the formula Pr_S . \Box and \Box_0 will be used as shorthand for $\Box_{\mathbf{PA}}$ and $\Box_{\mathbf{I}\Delta_0 + \mathbf{exp}}$ respectively. By \Box_x we denote the formula containing x as a free variable, and such that for $n > 0$, \Box_n (the result of substituting \bar{n} for x in \Box_x) is $\Box_{\mathbf{I}\Sigma_n}$. The symbol \Diamond_S is used for the dual of \Box_S , i.e. as an abbreviation for $\neg \Box_S \neg A$.

We use the dot notation as usual, thus $\Box_S A(\dot{x})$ means that the numeral for the value of x has been substituted for the free variable of the formula A inside \Box_S . If the intended meaning is clear from the context, we will often simply write $\Box_S A(x)$ instead of $\Box_S A(\dot{x})$. We recall that any theory S extending $\mathbf{I}\Delta_0 + \mathbf{exp}$ is provably Σ_1 -complete, meaning that for any Σ_1 -formula A ,

$$I\Delta_0+\mathbf{exp} \vdash A(x) \rightarrow \Box_S A(\dot{x}).$$

It is well-known that if S is as above, then the *Hilbert-Bernays-Löb derivability conditions* hold for \Box_S verifiably in $I\Delta_0+\mathbf{exp}$:

1. if $S \vdash A$, then $I\Delta_0+\mathbf{exp} \vdash \Box_S A$
2. $I\Delta_0+\mathbf{exp} \vdash \Box_S(A \rightarrow B) \rightarrow (\Box_S A \rightarrow \Box_S B)$
3. $I\Delta_0+\mathbf{exp} \vdash \Box_S A \rightarrow \Box_S \Box_S A$

In fact, 2 and 3 also hold with internal variables ranging over the sentences A and B . The following theorem is known as the Diagonal Lemma, or Gödel-Carnap Fixed Point Lemma.

Theorem 1 *Let A be a $\Sigma_n(\Pi_n)$ -formula whose free variables are exactly x_0, \dots, x_n . There is a $\Sigma_n(\Pi_n)$ -formula β with exactly the same free variables, and such that*

$$I\Delta_0+\mathbf{exp} \vdash \beta(x_1, \dots, x_n) \leftrightarrow A(\ulcorner \beta(v_1, \dots, v_n) \urcorner, x_1, \dots, x_n).$$

We say that $\beta(x_1, \dots, x_n)$ is a fixed point of $A(x_0 \dots, x_n)$.

Theorem 1, together with the Hilbert-Bernays-Löb derivability conditions for \Box_S , implies that Löb’s principle for \Box_S is verifiable in $I\Delta_0+\mathbf{exp}$ [2, Theorem 3.2]:

$$I\Delta_0+\mathbf{exp} \vdash \Box_S(\Box_S A \rightarrow A) \rightarrow \Box_S A.$$

Principles valid in the Gödel-Löb provability logic **GL** (see Sect. 2.3) can thus be used when reasoning about S in $I\Delta_0+\mathbf{exp}$.

Consider the sequence of theories $\{T_n\}_{n \in \omega}$, where $T_0 = I\Delta_0+\mathbf{exp}$, and for $n > 0$, $T_n = I\Sigma_n$. The basic facts concerning $\{T_n\}_{n \in \omega}$ are verifiable in $I\Delta_0+\mathbf{exp}$:

$$I\Delta_0+\mathbf{exp} \vdash \Box A \leftrightarrow \exists x \Box_x A \tag{2}$$

$$I\Delta_0+\mathbf{exp} \vdash \forall x, y (\Box_x A \wedge x < y \rightarrow \Box_y A) \tag{3}$$

Furthermore, $I\Delta_0+\mathbf{exp}$ verifies that each T_{n+1} proves *uniform Π_{n+2} -reflection* for T_n :

$$I\Delta_0+\mathbf{exp} \vdash \forall x (\Box_{x+1}(\forall A \in \Pi_{x+2} \rightarrow \forall y (\Box_x A(\dot{y}) \rightarrow A(y)))) \tag{4}$$

That (4) holds was first stated in [21]; an inspection of the proof shows that it can be verified in $I\Delta_0+\mathbf{exp}$. Throughout this paper, we shall refer to properties (3) and (4) as *monotonicity* and *reflection* respectively.

Oracles We recall that for $n > 1$, there is a partial truth definition tr_{Π_n} for Π_n -sentences in $I\Delta_0+\mathbf{exp}$ [7, Section V.5(b)]. The formula tr_{Π_n} is Π_n and satisfies Tarski’s conditions [7, Definition I.1.74].

Let S be an r.e. extension of $I\Delta_0+\mathbf{exp}$. Using tr_{Π_n} , one can define in a natural way the provability predicate $\Box_S^{\Pi_n}$ of S together with a Π_n -oracle. Thus $\Box_x^{\Pi_1}$ is an intensionally

correct provability predicate for $I\Sigma_x$ together with a Π_1 -oracle. Note that $\Box_x^{\Pi_1}$ is a Σ_2 -formula. It follows from (4) that $I\Delta_0+\text{exp} \vdash \forall x \Box_{x+1} \neg \Box_x \perp$. Inspection of the proof shows that it goes through when we add Π_1 -sentences that are known to be true to formal proofs, thus we have: $I\Delta_0+\text{exp} \vdash \forall x \Box_{x+1} \neg \Box_x^{\Pi_1} \perp$. Also this property will be referred to as *reflection*.

2.2 Interpretability

The notion of interpretability that we are interested in is that of *relative interpretability*, first introduced and carefully studied by Tarski, Mostowski and Robinson [29]. Since all theories considered here have pairing, it is safe to focus our attention on one-dimensional interpretations.

Definition 1 *Let S and T be first-order theories whose languages are \mathcal{L}_S and \mathcal{L}_T . An interpretation j of S in T is a tuple $\langle \delta, \tau \rangle$, where δ is an \mathcal{L}_T -formula with one free variable, and τ a mapping from relation symbols³ R of \mathcal{L}_S to formulas R^τ of \mathcal{L}_T , where the number of free variables of R^τ is equal to the arity of R . We extend τ to a translation from all formulas of \mathcal{L}_S to formulas of \mathcal{L}_T by requiring:*

- i. $(R(x_1, \dots, x_n))^\tau = R^\tau(x_1, \dots, x_n)$
- ii. $(A \rightarrow B)^\tau = A^\tau \rightarrow B^\tau$
- iii. $\perp^\tau = \perp$
- iv. $(\forall x A)^\tau = \forall x (\delta(x) \rightarrow A^\tau)$

Finally, we require that $T \vdash \exists x \delta(x)$, and $T \vdash A^\tau$ for all axioms A of S .

We write $j : T \triangleright S$ if j is an interpretation of S in T , and $T \triangleright S$ if $j : T \triangleright S$ for some j . We say that T and S are mutually interpretable, and write $T \equiv S$, if $T \triangleright S$ and $S \triangleright T$. In this paper we are concerned with interpretability between finite extensions of PA , i.e. theories of the form $\text{PA} + A$, where A is an \mathcal{L} -sentence. We write $A \triangleright B$ as an abbreviation for $\text{PA} + A \triangleright \text{PA} + B$.

Interpretability, like provability, is a syntactical notion, and can therefore be formalized in PA . We write $A \triangleright B$ for the arithmetical sentence expressing that $\text{PA} + A$ interprets $\text{PA} + B$.

The following theorem concerning interpretability over PA is referred to as the Orey-Hájek Characterization.

Theorem 2 ($I\Delta_0+\text{exp}$) *The following are equivalent:*

- i. $A \triangleright B$
- ii. For all n , $\text{PA} \vdash A \rightarrow \Diamond_n B$
- iii. For any $C \in \Pi_1$, $\text{PA} \vdash B \rightarrow C$ implies $\text{PA} \vdash A \rightarrow C$

Theorem 2 is implicit in [22], and was first explicitly stated in [6] and in [8]. Item iii was added in [5]. Inspection of the proof shows that it can be verified in $I\Delta_0+\text{exp}$.

³ We assume here that S is formulated in a purely relational way. This restriction is not essential – function symbols can be replaced by relation symbols by a well-known algorithm (for details, see [31, Section 7.3]).

2.3 Modal logic

We denote by \mathcal{L}_\square the language of propositional modal logic containing a unary modality \square , and by $\mathcal{L}_{\square\triangleright}$ the language \mathcal{L}_\square together with a binary modality \triangleright .

Provability logic The axiom schemata of the Gödel-Löb provability logic **GL** include all propositional tautologies in the language \mathcal{L}_\square , and furthermore:

$$(K) \quad \square(A \rightarrow B) \rightarrow (\square A \rightarrow \square B)$$

$$(L) \quad \square(\square A \rightarrow A) \rightarrow \square A$$

The inference rules of **GL** are modus ponens and necessitation:

$$\text{if } \mathbf{GL} \vdash A, \text{ then } \mathbf{GL} \vdash \square A.$$

We note that $\mathbf{GL} \vdash \square A \rightarrow \square \square A$ (see e.g. [2, p.11]).

The axiom schemata of the interpretability logic **ILM** include all propositional tautologies in the language $\mathcal{L}_{\square\triangleright}$, the axiom schemata of **GL**, and:

$$(J1) \quad \square(A \rightarrow B) \rightarrow A \triangleright B$$

$$(J2) \quad (A \triangleright B) \wedge (B \triangleright C) \rightarrow (A \triangleright C)$$

$$(J3) \quad (A \triangleright C) \wedge (B \triangleright C) \rightarrow (A \vee B) \triangleright C$$

$$(J4) \quad A \triangleright B \rightarrow (\diamond A \rightarrow \diamond B)$$

$$(J5) \quad \diamond A \triangleright A$$

$$(M) \quad A \triangleright B \rightarrow (A \wedge \square C) \triangleright (B \wedge \square C)$$

The rules of **ILM** are modus ponens, and necessitation for \square .

Definition 2 Let φ be an \mathcal{L} -formula with one free variable. A φ -realization is a function $*$ from the propositional letters of \mathcal{L}_\square to \mathcal{L} -sentences. The domain of $*$ is extended to all $\mathcal{L}_{\square\triangleright}$ -formulas by requiring that it commutes with the propositional connectives, and furthermore

$$(\square A)^* := \varphi(\ulcorner A^* \urcorner).$$

Given an \mathcal{L} -formulas φ and ψ with one and two free variables respectively, the notion of a φ - ψ -realization (a function from $\mathcal{L}_{\square\triangleright}$ -formulas to \mathcal{L} -sentences) is defined in a similar way.

Theorem 3 Let T be a r.e. and Σ_1 -sound extension of $\mathbf{I}\Delta_0 + \mathbf{exp}$, and let Pr_T be an intensionally correct provability predicate of T . Write \mathcal{R} for the set of all Pr_T -realizations. Then for all $A \in \mathcal{L}_\square$,

$$\mathbf{GL} \vdash A \quad \text{if and only if} \quad \text{for all } * \in \mathcal{R}, \mathbf{I}\Delta_0 + \mathbf{exp} \vdash A^*.$$

The direction from left to right follows by the fact that the Hilbert-Bernays-Löb derivability conditions for Pr_T are verifiable in $\mathbf{I}\Delta_0 + \mathbf{exp}$ [11, 18]. The proof of the other

direction with $T = \text{PA}$ is due to Solovay [27]. De Jongh et al. [13] extended the result to Σ_1 -sound theories containing $\text{I}\Delta_0 + \text{exp}$.

We recall that a theory is said to be *essentially reflexive* if it proves the consistency of each of its finite subtheories, and the same holds for every consistent extension in the same language. It is well-known that PA is essentially reflexive [7, Theorem III.2.35].

Theorem 4 *Let T be r.e., essentially reflexive, and Σ_1 -sound. Let Pr_T be an intensionally correct provability predicate of T , and Int an \mathcal{L} -formula (with two free variables) representing interpretability in T . Write \mathcal{R} for the set of all Pr_T - Int -realizations. For all $A \in \mathcal{L}_{\triangleright}$,*

$$\text{ILM} \vdash A \quad \text{if and only if} \quad \text{for all } * \in \mathcal{R}, T \vdash A^*.$$

Theorem 4 was proven independently by Shavrukov [24] and Berarducci [1].

3 Methodological considerations

Write $(\mathcal{V}, \triangleright)$ for the degrees of finite extensions of PA , ordered by the relation of interpretability. As mentioned in Sect. 1, it was proven by Švejdar that the structure $(\mathcal{V}, \triangleright)$ is a lattice. Our question is whether this fact is seen from the perspective of the interpretability logic ILM .

Indeed, the fact that $(\mathcal{V}, \triangleright)$ is a lower semilattice is reflected in the axioms of ILM . Principles J1 and J2 imply that the ordering given by \triangleright is reflexive and transitive, and furthermore that $A \triangleright A \vee B$ and $B \triangleright A \vee B$, i.e. that $[A \vee B]$ is a lower bound of $[A]$ and $[B]$ in $(\mathcal{V}, \triangleright)$. Principle J3 states that it is in fact the *greatest* lower bound, i.e. the infimum of $[A]$ and $[B]$.

To see that, in general, the supremum of $[A]$ and $[B]$ is not $[A \wedge B]$, it suffices to consider the so-called *Orey sentences* discovered by Orey in [22, Theorem 2.4]. An Orey sentence is any sentence O with $\top \triangleright O$ and $\top \triangleright \neg O$. An Orey sentence and its negation are thus both in $[\top]$, and hence also their supremum is an element of $[\top]$, which is clearly not the case for the sentence $O \wedge \neg O$.

Recall the structure $(\mathcal{D}, \triangleright)$ of the interpretability degrees of all r.e. extensions of PA . It follows from Theorem 2 that the supremum of $[\text{PA} + A]$ and $[\text{PA} + B]$ in $(\mathcal{D}, \triangleright)$ is $[S]$, where S is the following infinite theory:

$$S := \text{PA} + \{\diamond_n A \wedge \diamond_n B \mid n \in \omega\}.$$

Švejdar’s construction of suprema in $(\mathcal{V}, \triangleright)$ (see (1), or Theorem 5 in Sect. 4) can be seen as a way of compressing the information in S into a single sentence. This construction employs a language much more complicated than that available in the propositional system ILM . Indeed, as shown in Appendix A.1, the existence of interpretability suprema lies beyond the expressive power of ILM .

We want to enrich the language of ILM by adding to it a new binary modality \odot , together with the following axiom S , stating that \odot is an interpretability supremum:

$$(\text{S}) \quad (C \triangleright A) \wedge (C \triangleright B) \leftrightarrow (C \triangleright A \odot B).$$

We want to interpret \odot as an *arithmetical* formula. The following definition states what we are looking for.

Definition 3 An \mathcal{L} -formula σ with two free variables is a supremum implementation if for all \mathcal{L} -sentences A, B, C , we have that $\text{PA} \vdash (C \triangleright A) \wedge (C \triangleright B) \leftrightarrow C \triangleright \sigma(A, B)$.

Given a supremum implementation σ , we can extend the notion of an arithmetical realization to include formulas of the extended modal language:

$$(A \odot B)^* := \sigma(\ulcorner A^* \urcorner, \ulcorner B^* \urcorner). \tag{5}$$

Let \mathcal{R} be the set of all realizations where the ILM part is as in Theorem 4, and \odot is interpreted as in (5). Denote by ILMS the system ILM together with axiom S. Then it is clear that for all modal formulas A ,

$$\text{If ILMS} \vdash A, \text{ then for all } * \in \mathcal{R}, \text{ PA} \vdash A^*.$$

In other words ILMS is arithmetically sound. For arithmetical completeness, we need to make sure that all propositional schemata involving σ that are theorems of PA are also provable in ILMS. However, it turns out that what exactly is provable in PA about a supremum implementation is far from determined by the fact that the latter satisfies Definition 3. For example, while for any such σ it is clear that $\sigma(A, B)$ and $\sigma(B, A)$ are, verifiably in PA, mutually interpretable, there is no a priori reason why

$$\sigma(A, B) \leftrightarrow \sigma(B, A) \tag{6}$$

should be provable in PA, or even true (see Appendix A.2). On the other hand, the supremum implementations we encounter in the following sections all satisfy (6).

Given the above, should we add $A \odot B \leftrightarrow B \odot A$ to the axioms of ILMS? The answer depends on which supremum implementation(s) we have in mind. In contrast to formalized provability and interpretability, there is no strong intuition as to what constitutes a *natural* supremum implementation. Thus our choice will depend on practical and esthetical criteria. For example, we prefer implementations that allow for a nice Kripke semantics. The following definition states some nice features that a supremum implementation could have.

Definition 4 Let A, A', B, B' range over all \mathcal{L} -sentences, and let σ be an \mathcal{L} -formula with two free variables. We say that σ is extensional if

$$\text{PA} \vdash A \leftrightarrow A' \text{ and } \text{PA} \vdash B \leftrightarrow B' \quad \Rightarrow \quad \text{PA} \vdash \sigma(A, B) \leftrightarrow \sigma(A', B'),$$

and σ is provably extensional if the above is verifiable in PA, i.e.

$$\text{PA} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \rightarrow \Box(\sigma(A, B) \leftrightarrow \sigma(A', B')).$$

We say that σ is monotone if

$$\text{PA} \vdash A \rightarrow A' \text{ and } \text{PA} \vdash B \rightarrow B' \quad \Rightarrow \quad \text{PA} \vdash \sigma(A, B) \rightarrow \sigma(A', B'),$$

and provably monotone if the above is verifiable in PA, i.e.

$$\text{PA} \vdash \Box(A \rightarrow A') \wedge \Box(B \rightarrow B') \rightarrow \Box(\sigma(A, B) \rightarrow \sigma(A', B')).$$

We note that the above properties are ordered by strength, with extensionality being the weakest.

We conclude this section by considering a supremum implementation property that is so nice that we did not even think of it before Shavrukov gave us an example.

Definition 5 A supremum implementation σ is adapted if there is an \mathcal{L} -formula τ containing one free variable, and such that for all \mathcal{L} -sentences A and B ,

$$\sigma(A, B) = \tau(A) \wedge \tau(B).$$

We call a formula τ as in Definition 5 a *supremum adapter*. A supremum adapter is thus a *unary* operator that, in combination with conjunction, provides us with a supremum implementation.⁴

4 Varieties of supremum implementations

This section studies the supremum implementation based on Švejdar’s argument showing that $(\mathcal{V}, \triangleright)$ is a lattice, as well as a dual construction discovered by Visser. These implementations are given as fixed points of certain arithmetical formulas. We prove the existence of unique explicit fixed points for these formulas. The purpose of this section is mainly methodological; it can be safely skipped without affecting insight into the rest of the article.

4.1 Definition and verification of the target property

As mentioned above, it follows from Theorem 2 that the *infinite* theory

$$\text{PA} + \{\Diamond_n A \wedge \Diamond_n B \mid n \in \omega\} \tag{7}$$

is an interpretability supremum of $\text{PA} + A$ and $\text{PA} + B$. The idea of Švejdar’s construction is to compress the information contained in (7) into a single sentence.

Theorem 5 ([28, Theorem 4.4]) *By Theorem 1, let ϑ be such that*

$$\text{PA} \vdash \vartheta \leftrightarrow \forall x (\Diamond_x \vartheta \rightarrow \Diamond_x A \wedge \Diamond_x B).$$

Then ϑ is (verifiably in $\text{I}\Delta_0 + \text{exp}$) a supremum of A and B in $(\mathcal{V}, \triangleright)$.

We note that the complexity of ϑ is Π_2 . A dual construction, yielding a Σ_2 -supremum, was discovered by Visser.

⁴ We think of τ as *adapting* A and B , so that their conjunction can be used to get the supremum.

Theorem 6 *By Theorem 1, let ϑ be such that*

$$\text{PA} \vdash \vartheta \leftrightarrow \exists x (\Box_x \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B)).$$

Then ϑ is (verifiably in $\text{I}\Delta_0 + \text{exp}$) a supremum of A and B in $(\mathcal{V}, \triangleright)$.

Proof Argue in $\text{I}\Delta_0 + \text{exp}$. We show that

- i. $\vartheta \triangleright A$ and $\vartheta \triangleright B$
- ii. Whenever $C \triangleright A$ and $C \triangleright B$, also $C \triangleright \vartheta$

We first show i. By Theorem 2, it suffices to show that ϑ implies $\Diamond_n A$ and $\Diamond_n B$ for any standard n . Assume ϑ . Thus there is some a with $\Box_a \neg \vartheta$, $\Diamond_a A$, and $\Diamond_a B$. By essential reflexivity, a must be greater than any standard n . By monotonicity, we thus have $\Diamond_n A$ and $\Diamond_n B$ for any such n .

For ii, let C be such that $C \triangleright A$ and $C \triangleright B$. We claim that $C \wedge \neg \vartheta \triangleright \vartheta$. Note that $\neg \vartheta$ is the sentence $\forall x ((\Diamond_x A \wedge \Diamond_x B) \rightarrow \Diamond_x \vartheta)$. Since $C \triangleright A$ and $C \triangleright B$, by Theorem 2 we have that C implies $\Diamond_n A$ and $\Diamond_n B$ for any standard n . But then $C \wedge \neg \vartheta$ implies $\Diamond_n \vartheta$ for any such n , whence $C \wedge \neg \vartheta \triangleright \vartheta$ again by Theorem 2. Finally, note that clearly $C \wedge \vartheta \triangleright \vartheta$. By axiom J3 of ILM, we therefore have $(C \wedge \vartheta) \vee (C \wedge \neg \vartheta) \triangleright \vartheta$, i.e. $C \triangleright \vartheta$. □

Strictly speaking, theorems 5 and 6 do not yet provide us with supremum implementations in the sense of Definition 3. Corresponding to Theorem 5, we would want a formula σ with two free variables, such that for any sentences A and B ,

$$\text{PA} \vdash \sigma(A, B) \leftrightarrow \forall x (\Diamond_x \sigma(A, B) \rightarrow \Diamond_x A \wedge \Diamond_x B). \tag{8}$$

Such a formula can be found by using Theorem 1 (see Appendix B.1). Having this in mind, we shall from now on say Švejdar’s supremum (implementation) to refer to a formula σ as in (8); similarly for Visser’s supremum.

4.2 Existence of unique explicit fixed points

The suprema introduced above are given as fixed points of certain formulas. Before proving the existence of unique and explicit fixed points for these formulas, we examine a surprising feature of Švejdar’s construction. Švejdar’s supremum of A and B is obtained as a fixed point of the following formula:⁵

$$\forall x (\Diamond_x Y \rightarrow (\Diamond_x A \wedge \Diamond_x B)).$$

In particular the supremum of \perp and \perp is given as a fixed point of

$$\forall x (\Diamond_x Y \rightarrow \Diamond_x \perp). \tag{9}$$

⁵ The capital variable Y indicates that we are interested in fixed points with respect to this variable.

Example 1 Write $\Psi_{\perp}(Y)$ for the formula in (9). We have the following:

- i. If $\text{PA} \vdash \vartheta \leftrightarrow \Psi_{\perp}(\vartheta)$, then $\text{PA} \vdash \vartheta \leftrightarrow \perp$
- ii. $\text{PA} \not\vdash \perp \leftrightarrow \Psi_{\perp}(\perp)$

In other words every fixed point of $\Psi_{\perp}(Y)$ is equivalent to \perp , however \perp itself is not a fixed point of $\Psi_{\perp}(Y)$. To see that i holds, suppose that $\text{PA} \vdash \vartheta \leftrightarrow \Psi_{\perp}(\vartheta)$, i.e.

$$\text{PA} \vdash \vartheta \leftrightarrow \forall x (\Box_x \top \rightarrow \Box_x \neg \vartheta).$$

Since $\text{PA} \vdash \forall x \Box_x \top$, it follows that $\text{PA} \vdash \vartheta \leftrightarrow \Box_0 \neg \vartheta$. Since $\text{PA} \vdash \Box_0 \neg \vartheta \rightarrow \neg \vartheta$ by reflection, we thus have $\text{PA} \vdash \vartheta \leftrightarrow \perp$. For ii note that $\Psi_{\perp}(\perp)$ is $\forall x (\Diamond_x \perp \rightarrow \Diamond_x \perp)$, i.e. it is a tautology and thus clearly not provably equivalent to \perp .

Example 1 demonstrates that the formula $\Psi_{\perp}(Y)$ is not extensional, i.e. it is *not* the case that for all sentences A and B ,

$$\text{PA} \vdash A \leftrightarrow B \Rightarrow \text{PA} \vdash \Psi_{\perp}(A) \leftrightarrow \Psi_{\perp}(B). \tag{10}$$

Indeed, if ϑ is a fixed point of $\Psi_{\perp}(Y)$, then as shown above $\text{PA} \vdash \vartheta \leftrightarrow \perp$. On the other hand $\Psi_{\perp}(\perp)$ is a tautology and therefore not provably equivalent to $\Psi_{\perp}(\vartheta)$ – since the latter is equivalent to ϑ and hence to \perp .

We thus cannot apply to $\forall x (\Diamond_x Y \rightarrow (\Diamond_x A \wedge \Diamond_x B))$ Smoryński’s method [26] for establishing the uniqueness of fixed points of arithmetical formulas.⁶ The following theorem, due to Shavrukov, shows that unique explicit fixed points nevertheless exist.

Theorem 7 *Let $\varepsilon(A, B)$ be the formula*

$$(\Diamond_0 \top \rightarrow \Diamond_0 A \wedge \Diamond_0 B) \wedge \forall x (\Diamond_{x+1}(\Diamond_x A \wedge \Diamond_x B) \rightarrow \Diamond_{x+1} A \wedge \Diamond_{x+1} B),$$

where A and B are regarded as internal variables ranging over \mathcal{L} -sentences. Then

- i. $\text{PA} \vdash \varepsilon(A, B) \leftrightarrow \forall x (\Diamond_x \varepsilon(A, B) \rightarrow \Diamond_x A \wedge \Diamond_x B)$
- ii. For any ϑ , if $\text{PA} \vdash \vartheta \leftrightarrow \forall x (\Diamond_x \vartheta \rightarrow \Diamond_x A \wedge \Diamond_x B)$, then $\text{PA} \vdash \vartheta \leftrightarrow \varepsilon(A, B)$

Proof See Appendix B.2. □

A similar result holds for Visser’s supremum implementation.

Theorem 8 *Let $\varepsilon(A, B)$ be the formula*

$$\exists x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B))$$

where A and B are regarded as internal variables ranging over \mathcal{L} -sentences. Then

- i. $\text{PA} \vdash \varepsilon(A, B) \leftrightarrow \exists x (\Box_x \neg \varepsilon(A, B) \wedge (\Diamond_x A \wedge \Diamond_x B))$

⁶ Smoryński’s method for showing the uniqueness of fixed points of an arithmetical formula $\Psi(Y)$ relies on $\Psi(Y)$ being *PA-substitutable*, meaning that $\text{PA} \vdash \Box(A \leftrightarrow B) \rightarrow (\Psi(A) \leftrightarrow \Psi(B))$ for all A and B . Note that if $\Psi(Y)$ is *PA-substitutable*, then it is also extensional in the sense of Definition 4.

ii. For any ϑ , if $\text{PA} \vdash \vartheta \leftrightarrow \exists x (\Box \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$, then $\text{PA} \vdash \vartheta \leftrightarrow \varepsilon(A, B)$

Proof See Appendix B.2. □

It follows that any formula ε as in Theorem 7 or in Theorem 8 is a supremum implementation (using items i. and the proofs of theorems 5 and 6). With the explicit versions at hand, we can learn a bit more about Švejdar’s and Visser’s suprema.

Theorem 9 Švejdar’s and Visser’s supremum are both extensional.

Proof See Appendix B.3. □

Theorem 10 Visser’s supremum is not monotone.

Proof See Appendix B.4. □

Question 1 Is Švejdar’s supremum monotone?

5 Supremum adapters

The idea and the first examples of supremum adapters are due to Shavrukov. The versions introduced here were obtained by analyzing and simplifying the latter. Define

$$\begin{aligned} \sigma(A) &:= \forall x (\Diamond_x^{\Pi_1} \top \rightarrow \Diamond_{x+1} A) \\ \tau(A) &:= \exists x (\Box_x^{\Pi_1} \perp \wedge \Diamond_x A) \end{aligned}$$

where A is regarded as an internal variable ranging over sentences. Write ∇A for either $\sigma(A)$ or $\tau(A)$. In this section, we show that ∇ is a supremum adapter, i.e. that for all \mathcal{L} -sentences A, B , and C , it is provable in PA that

$$(C \triangleright A) \wedge (C \triangleright B) \leftrightarrow C \triangleright (\nabla A \wedge \nabla B).$$

In fact, we show that the above is already verifiable in $\text{I}\Delta_0 + \text{exp}$.

Lemma 1 ($\text{I}\Delta_0 + \text{exp}$) For all n , $\text{PA} \vdash \nabla A \rightarrow \Diamond_n A$.

Proof Straightforward from the definition of σ, τ , and the fact that $\text{I}\Sigma_{n+1} \vdash \Diamond_n^{\Pi_1} \top$ for all n by reflection. □

Lemma 2 $\text{I}\Delta_0 + \text{exp} \vdash \forall z (\Box_z (\neg \nabla A \vee \neg \nabla B) \rightarrow \Box_z (\Box_z \neg A \vee \Box_z \neg B))$.

Proof Argue $\text{I}\Delta_0 + \text{exp}$, assuming

$$\Box_z (\neg \nabla A \vee \neg \nabla B). \tag{11}$$

We first show that (11) implies

$$\Box_z (\Box_z^{\Pi_1} \perp \rightarrow (\Box_z \neg A \vee \Box_z \neg B)). \tag{12}$$

Suppose first that (11) is $\Box_z(\neg\sigma(A) \vee \neg\sigma(B))$, i.e.

$$\Box_z(\exists x (\Diamond_x^{\Pi_1} \top \wedge \Box_{x+1} \neg A) \vee \exists x (\Diamond_x^{\Pi_1} \top \wedge \Box_{x+1} \neg B)). \tag{13}$$

It follows that

$$\Box_z \exists x (\Diamond_{x-1}^{\Pi_1} \top \wedge (\Box_x \neg A \vee \Box_x \neg B)). \tag{14}$$

We reason in \Box_z . By (14), let y be such that $\neg\Box_y^{\Pi_1} \perp \wedge (\Box_{y+1} \neg A \vee \Box_{y+1} \neg B)$. Assuming $\Box_z^{\Pi_1} \perp$, we have $y < z$, whence by monotonicity $\Box_z \neg A \vee \Box_z \neg B$. We exit \Box_z . Note that we have shown (12). The other possibility is that (11) is $\Box_z(\neg\tau(A) \vee \neg\tau(B))$, i.e.

$$\Box_z(\forall x (\Box_x^{\Pi_1} \perp \rightarrow \Box_x \neg A) \vee \forall x (\Box_x^{\Pi_1} \perp \rightarrow \Box_x \neg B)). \tag{15}$$

From this it follows that $\Box_z \forall x (\Box_x^{\Pi_1} \perp \rightarrow (\Box_x \neg A \vee \Box_x \neg B))$, whence clearly also (12). Thus we have (12) whenever (11) holds. We now show that (12) implies

$$\Box_z(\Box_z^{\Pi_1} (\Box_z \neg A \vee \Box_z \neg B) \rightarrow \Box_z \neg A \vee \Box_z \neg B). \tag{16}$$

To see that (16) holds, reason inside \Box_z , assuming

$$\Box_z^{\Pi_1} (\Box_z \neg A \vee \Box_z \neg B). \tag{17}$$

If $\Box_z \neg A \vee \Box_z \neg B$, we are done. So assume that not: $\Box_z \neg A \vee \Box_z \neg B$, i.e.

$$\Diamond_z A \wedge \Diamond_z B. \tag{18}$$

Since (18) is a Π_1 -sentence, we have $\Box_z^{\Pi_1} (\Diamond_z A \wedge \Diamond_z B)$ and thus $\Box_z^{\Pi_1} \perp$ in combination with (17). The desired conclusion $\Box_z \neg A \vee \Box_z \neg B$ now follows by (12). Exiting the world inside \Box_z , we have established (16). From the latter it clearly follows that

$$\Box_z(\Box_z(\Box_z \neg A \vee \Box_z \neg B) \rightarrow \Box_z \neg A \vee \Box_z \neg B), \tag{19}$$

and thus $\Box_z(\Box_z \neg A \vee \Box_z \neg B)$ by Löb’s principle for $\mathbf{I}\Sigma_z$.

Lemma 3 ($\mathbf{I}\Delta_0 + \mathbf{exp}$) *For all n , $\mathbf{PA} \vdash \Diamond_n A \wedge \Diamond_n B \rightarrow \Diamond_n(\nabla A \wedge \nabla B)$.*

Proof Argue in $\mathbf{I}\Delta_0 + \mathbf{exp}$. By contraposition, it suffices to show that for all n ,

$$\mathbf{I}\Sigma_{n+1} \vdash \Box_n(\neg\nabla A \vee \neg\nabla B) \rightarrow \Box_n \neg A \vee \Box_n \neg B.$$

Fix an n . From Lemma 2 it follows that

$$\mathbf{I}\Sigma_{n+1} \vdash \Box_n(\neg\nabla A \vee \neg\nabla B) \rightarrow \Box_n(\Box_n \neg A \vee \Box_n \neg B).$$

Since $\mathbf{I}\Sigma_{n+1} \vdash \Box_n(\Box_n \neg A \vee \Box_n \neg B) \rightarrow (\Box_n \neg A \vee \Box_n \neg B)$ by reflection, we are done. □

Theorem 11 ($I\Delta_0+\text{exp}$) *For all \mathcal{L} -sentences $A, B,$ and $C,$*

$$(C \triangleright A) \wedge (C \triangleright B) \leftrightarrow C \triangleright (\nabla A \wedge \nabla B).$$

Proof Straightforward consequence of lemmas 1, 3, and Theorem 2. □

6 Nonstandard provability predicates

The supremum adapters introduced in the previous section can be seen as consistency statements corresponding to certain (nonstandard) provability predicates. We shall now investigate the consequences of this perspective, showing in particular that these provability predicates satisfy the principles of the provability logic **GL**. In Sect. 6.3, we establish some principles for the bimodal provability logic of such a nonstandard notion of provability, together with ordinary provability.

We use modal notation, writing ∇A and $\triangleright A$ for the sentences $\sigma(A)$ and $\tau(A)$ as in the previous section. The provability predicates \blacktriangle and \triangle are defined as their duals, i.e. $\blacktriangle A := \neg\nabla\neg A$ and $\triangle A := \neg\triangleright\neg A$. Spelling this out:

$$\begin{aligned} \blacktriangle A &= \exists x (\Box_{x+1} A \wedge \Diamond_x^{\Pi_1} \top) \\ \triangle A &= \forall x (\Box_x^{\Pi_1} \perp \rightarrow \Box_x A) \end{aligned}$$

Thus an \mathcal{L} -formula A is \blacktriangle -provable just in case there is some n such that A is provable (in the usual sense) in $I\Sigma_{n+1}$, and the theory $I\Sigma_n$ together with a Π_1 -oracle is consistent. Note that since for all n it is true (in the standard model) that $I\Sigma_n$ together with a Π_1 -oracle is consistent, the extension of \blacktriangle in the standard model coincides with ordinary **PA**-provability defined by \Box . However this might not be the case in a model where **PA** together with a Π_1 -oracle is inconsistent. This is why we say that \blacktriangle is a *nonstandard* provability predicate for **PA**. Note also that \blacktriangle is a Σ_3 -formula.

As for the other triangle, we note that a formula A is \triangle -provable if for all n , a proof of inconsistency of $I\Sigma_n$ together with a Π_1 -oracle implies the existence of a usual $I\Sigma_n$ -proof of A . We note that \triangle is a Π_2 -formula.

6.1 Relating the two triangles

Arguing in **PA**, let $\mu + 1$ be the smallest number such that $I\Sigma_{\mu+1}$ together with a Π_1 -oracle is inconsistent. It is easy to see that if $\mu + 1$ exists, then

$$\blacktriangle A \leftrightarrow \triangle A \leftrightarrow \Box_{\mu+1} A$$

If $\mu + 1$ does not exist, i.e. if **PA** together with a Π_1 -oracle is consistent, then $\blacktriangle A$ is equivalent to $\Box A$, while $\triangle A$ is equivalent to \top . Both \blacktriangle and \triangle are therefore equivalent to $\Box_{\mu+1}$, the only difference being in how the latter is interpreted in case $\mu + 1$ does not exist. This leads to the following alternative definition of \blacktriangle and \triangle :

$$\blacktriangle A := \begin{cases} \Box_{\mu+1} A & \text{if } \Box^{\Pi_1} \perp \\ \Box A & \text{otherwise} \end{cases} \tag{20}$$

$$\triangle A := \begin{cases} \Box_{\mu+1} A & \text{if } \Box^{\Pi_1} \perp \\ \top & \text{otherwise} \end{cases} \tag{21}$$

Using the above observations, we see that the two triangles are closely related.

Lemma 4 *i.* $\text{PA} \vdash \blacktriangle A \leftrightarrow (\Diamond^{\Pi_1} \top \wedge \Box A) \vee (\Box^{\Pi_1} \perp \wedge \triangle A)$
ii. $\text{PA} \vdash \triangle A \leftrightarrow (\blacktriangle A \vee \Diamond^{\Pi_1} \top)$

Proof For *i*, reason in **PA**. First suppose $\blacktriangle A$. From the definition of \blacktriangle , it follows that also $\Box A$. If $\Diamond^{\Pi_1} \top$, we are done. If $\Box^{\Pi_1} \perp$, then $\blacktriangle A$ and $\triangle A$ are equivalent, whence $\triangle A$. For the other direction, suppose that $\Diamond^{\Pi_1} \top$ and $\Box A$. Then $\blacktriangle A$ follows straight from the definition. And if $\Box^{\Pi_1} \perp$, then $\triangle A$ if and only of $\blacktriangle A$. The proof of *ii* is similar. □

Given existential sentences $A = \exists x A'$ and $B = \exists y B'$, we write $A \leq B$ for the sentence $\exists x (A' \wedge \forall y < x \neg B')$. Thus $A \leq B$ says that the formula A has a witness, and this witness is smaller than or equal to the smallest witness of B . Similarly, we write $A < B$ for the sentence $\exists x (A' \wedge \forall y \leq x \neg B')$. If C is of the form $A \leq B$ we write C^\perp for $B < A$, and if C is of the form $A < B$, we write C^\perp for $B \leq A$. Note that C^\perp implies $\neg C$, but not the other way around. Thus C^\perp can be seen as a strong negation of C . We write $\Box_* A$ for $\exists x \Box_x A$, and similarly $\Box_*^{\Pi_1} A$ for $\exists x \Box_x^{\Pi_1} A$.

It is easy to see that

$$\text{PA} \vdash \blacktriangle A \leftrightarrow \Box_* A \leq \Box_*^{\Pi_1} \perp.$$

Now, $\blacktriangle^\perp A$ is the sentence $\Box_*^{\Pi_1} \perp < \Box_* A$. Rewriting the definition of \triangle according to the witness comparison notation, we see that \triangle is equivalent the negation of the latter:

$$\text{PA} \vdash \triangle A \leftrightarrow \neg(\Box_*^{\Pi_1} \perp < \Box_* A).$$

Thus $\text{PA} \vdash \triangle A \leftrightarrow \neg \blacktriangle^\perp A$, whence \triangle is, in a sense, the double negation of \blacktriangle .

6.2 Provability logic of \triangle

We use the symbol \triangle to refer to either \blacktriangle or \triangle . We show that $\text{I}\Delta_0 + \text{exp}$ verifies the principles of **GL** for \triangle .

Theorem 12 *1.* If $\text{PA} \vdash A$, then $\text{PA} \vdash \triangle A$
2. $\text{I}\Delta_0 + \text{exp} \vdash \triangle(A \rightarrow B) \rightarrow (\triangle A \rightarrow \triangle B)$
3. $\text{I}\Delta_0 + \text{exp} \vdash \triangle A \rightarrow \triangle \triangle A$

Proof Item 1 follows by reflection, and item 2 by principle **K** for $\text{I}\Sigma_x$. For 3, we first consider $\blacktriangle A$. Argue in $\text{I}\Delta_0 + \text{exp}$, assuming $\exists x (\Box_{x+1} A \wedge \neg \Box_x^{\Pi_1} \perp)$. We want to show

$$\exists x (\Box_{x+1} \exists y (\Box_{y+1} A \wedge \neg \Box_y^{\Pi_1} \perp) \wedge \neg \Box_x^{\Pi_1} \perp). \tag{22}$$

Let x be such that $\Box_{x+1}A$ and $\neg\Box_x^{\Pi_1}\perp$. We have $\Box_{x+1}\Box_{x+1}A$ by **GL** for $I\Sigma_x$, and $\Box_{x+1}\neg\Box_x^{\Pi_1}\perp$ by reflection. Therefore $\Box_{x+1}(\Box_{x+1}A \wedge \neg\Box_x^{\Pi_1}\perp)$, and thus clearly also (22). It remains to consider ΔA . Assuming $\forall x (\Box_x^{\Pi_1}\perp \rightarrow \Box_x A)$, we want to show:

$$\forall x (\Box_x^{\Pi_1}\perp \rightarrow \Box_x \forall y (\Box_y^{\Pi_1}\perp \rightarrow \Box_y A)).$$

Suppose $\Box_x^{\Pi_1}\perp$. We get $\Box_x A$ by assumption, and thus also $\Box_x\Box_x A$ and $\Box_x \forall y \geq x \Box_y A$. On the other hand, we have by reflection $\Box_x \forall y < x \neg\Box_y^{\Pi_1}\perp$. Combining the above yields $\Box_x \forall y (\Box_y^{\Pi_1}\perp \rightarrow \Box_y A)$ as required. \square

Remark 1 By examining the proof, it is clear that Theorem 12 also holds with internal quantifiers in items 2 and 3.

As explained in Sect. 2.1, it follows from Theorems 12 and 1 that

$$PA \vdash \Delta(\Delta A \rightarrow A) \rightarrow \Delta A,$$

and thus that **GL** is arithmetically sound with respect to Δ . We note that since Δ is not a Σ_1 -formula, arithmetical completeness does not follow by the usual method. However, as has been shown by Shavrukov and Henk, **GL** is nevertheless arithmetically complete with respect to \blacktriangle (see the forthcoming preprint [10]).

Question 2 Is **GL** arithmetically complete with respect to Δ ?

6.3 Joint provability logic of Δ and \square

We establish some principles for the joint provability logic of a supremum adapter provability, together with ordinary provability. As before, we use the symbol Δ to refer to either \blacktriangle or Δ .

Lemma 5 $I\Delta_0+\text{exp} \vdash \Box A \rightarrow \Delta\Box A$.

Proof It is easy to see that $S \rightarrow \Delta S$ holds (verifiably in $I\Delta_0+\text{exp}$) for any Σ_1 -formula S . The desired result follows by noting that $\Box A$ is a Σ_1 -formula. \square

Lemma 6 $I\Delta_0+\text{exp} \vdash \Box(\Box B \rightarrow A) \leftrightarrow \Box(\Box B \rightarrow \Delta A)$

Proof Argue in $I\Delta_0+\text{exp}$. By Theorem 11, $\neg A \equiv \nabla\neg A$, whence by Theorem 2, $\neg A$ is (provably in **PA**) Π_1 -conservative over $\nabla\neg A$ and vice versa. Since $\Box B$ is Π_1 , this means that $\Box(\neg A \rightarrow \Box B) \leftrightarrow \Box(\nabla\neg A \rightarrow \Box B)$. The desired result follows by contraposition. \square

Taking \top for B in Lemma 6, it follows that $I\Delta_0+\text{exp} \vdash \Box A \leftrightarrow \Delta\Box A$.

Monotonicity Returning to the business of supremum implementations, we use the above lemmas to show that supremum implementations based on our supremum adapters are monotone.

Lemma 7 $\vdash \Box(A \rightarrow B) \rightarrow \Box(\nabla A \rightarrow \nabla B)$.

Proof Argue in PA. If $\Box(A \rightarrow B)$, then $\Box(\neg B \rightarrow \neg A)$, and thus $\Box\Delta(\neg B \rightarrow \neg A)$ by the remark under Lemma 6. By the principles of GL for Δ and for \Box , this implies $\Box(\Delta\neg B \rightarrow \Delta\neg A)$, which in turn implies $\Box(\nabla A \rightarrow \nabla B)$. \square

6.4 Joint provability logic of \blacktriangle and \Box

We consider the joint provability logic of \blacktriangle and \Box . There are several reasons that make \blacktriangle a preferred object of study compared to Δ . As explained above, in contrast to Δ the formula \blacktriangle is an *extensionally correct* provability predicate for PA (albeit a nonstandard one). Second, we know that GL is arithmetically complete with respect to \blacktriangle , while we do not yet know whether the same holds for Δ . Finally, as is easily seen from the definition, we have that $\blacktriangle A \rightarrow \Box A$. On the other hand, there seems to be no modal principle relating Δ and \Box in such a simple way.

Remark 2 We note the similarity of \blacktriangle to the Feferman provability predicate:

$$\Box_f A := \exists x (\Box_x A \wedge \Diamond_x \top)$$

As shown in [4, Theorem 5.9], \Box_f is an extensionally correct provability predicate for PA, however at the same time $\text{PA} \vdash \Diamond_f \top$. The existence of \Box_f illustrates the need for a more careful formulation of Gödel’s Second Incompleteness Theorem than what is commonly stated. The Feferman provability predicate \Box_f has been further studied in [19,30], and [25].

Recall that $\blacktriangle A = \exists x (\Box_{x+1} A \wedge \Diamond_x^{\Pi_1} \top)$. Thus \blacktriangle differs from \Box_f by a Π_1 -oracle, and by a “+1”. Removing the Π_1 -oracle, we obtain the formula $\exists x (\Box_{x+1} A \wedge \Diamond_x \top)$ that is easily seen to be equivalent to $\Box A$. Remove the “+1” we obtain:

$$\Delta_f A := \exists x (\Box_x A \wedge \Diamond_x^{\Pi_1} \top). \tag{23}$$

Inspecting the proof of Theorem 11, it is clear that everything works when taking the dual ∇_f of Δ_f for ∇ . Thus ∇_f is also a supremum adapter.

On the other hand, the modal principles valid for Δ_f are rather different from those valid for \blacktriangle . It is easy to see that $\text{PA} \vdash \nabla_f \top$, from which it follows that Δ_f cannot satisfy the principles of GL. Indeed, examining the proof of Theorem 12, the “+1” in the definition of \blacktriangle seems to be essential for establishing transitivity ($\Delta A \rightarrow \Delta\Delta A$).

There is no a priori reason for preferring a supremum implementation based on \blacktriangle to one based on Δ_f . The advantage of the GL-satisfying version is that it satisfies a very well studied modal logic. As we will now see, also the joint provability logic of \blacktriangle and \Box has a relative in the existing literature.

Denote by $\mathcal{L}_{\Box\Delta}$ the language of propositional modal logic containing two unary modalities \Box and Δ .

Definition 6 *The axiom schemata of GLS include all propositional tautologies in the language $\mathcal{L}_{\square, \Delta}$, the axiom schemata of GL for both \square and Δ , and furthermore*

- (T1) $\Delta A \rightarrow \square A$
- (T2) $\square A \rightarrow \Delta \square A$
- (T3) $\square A \rightarrow \square \Delta A$
- (T4) $\square(\square B \rightarrow \Delta A) \rightarrow \square(\square B \rightarrow A)$

We note that (T3) and (T2) imply $\square(\square B \rightarrow A) \rightarrow \square(\square B \rightarrow \Delta A)$. In contrast, (T4) does not follow from the weaker $\square \Delta A \rightarrow \square A$. We denote by GLT the system GLS, with axiom (T4) replaced by $\square \Delta A \rightarrow \square A$.

Question 3 Is GLS arithmetically complete with respect to \square and \blacktriangle ?

The system GLT was studied by Lindström [17] due to its relation to Parikh's rule. The latter allows us to infer A , given that we have established the provability of A . It is thus a sort of reflection rule. Since Parikh's rule is admissible in PA, adding it to PA does not yield new theorems. As shown in [23], it does yield speed-up, meaning that some theorems have much shorter proofs when Parikh's rule is allowed. The equivalence of Parikh provability and ordinary provability is however not verifiable in PA. Writing \square_p for the formula representing Parikh provability, we have

$$\text{PA} \vdash \square_p A \leftrightarrow \exists x \square^x A,$$

where \square^x denotes the x -times iterated version of \square . Using the above, it is not difficult to see that the principles of GLT are valid when interpreting the modal symbol Δ as ordinary provability, and \square as Parikh provability. Indeed, it was proven by Lindström in [17] that GLT is arithmetically complete with respect to this interpretation. In [9], it is shown that GLT is also the arithmetically complete joint provability logic of slow and ordinary provability.

We note that both of the above interpretations of GLT are with respect to a pair of provability predicates where one is the ordinary provability predicate, and the other a *nonstandard* Σ_1 provability predicate for PA. Our conjecture for the system GLS on the other hand involves the ordinary provability predicate, together with a *nonstandard* Σ_3 -predicate. Dealing with the complexity of \blacktriangle while trying to answer Question 3 remains a challenge for future work.

Acknowledgements We wish to thank Volodya Shavrukov for several insights that had a profound effect on the evolution of the subject matter treated in this article. We are grateful to Dick de Jongh for useful discussions on the subject matter, including earlier versions of the article. We thank the anonymous referee for helpful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

A Methodological considerations

A.1 Undefinability of suprema in ILM

We show here that the existence of suprema in $(\mathcal{V}, \triangleright)$ is not expressible in the modal system ILM.

Lemma 8 *Let A be an $\mathcal{L}_{\square\triangleright}$ -formula. Then there exists a formula A^p containing no occurrences of \square and \triangleright , and such that*

$$\text{ILM} \vdash \square\perp \rightarrow (A \leftrightarrow A^p)$$

Proof By induction on the complexity of A . The base case and the propositional cases are trivial. If $A = \square B$ or $A = B \triangleright C$ for some B, C , then we can take \top for A^p (this is easily seen by using soundness and completeness of ILM with respect to Veltman frames, see [14]). □

Theorem 13 *There is no $\mathcal{L}_{\square\triangleright}$ -formula S , such that for all $\mathcal{L}_{\square\triangleright}$ -formulas A, B , and C ,*

$$\text{ILM} \vdash (C \triangleright A) \wedge (C \triangleright B) \leftrightarrow C \triangleright S(A, B).$$

Proof Suppose that such a formula S exists. Then in particular

$$\text{ILM} \vdash (\top \triangleright A) \wedge (\top \triangleright B) \leftrightarrow \top \triangleright S(A, B). \tag{24}$$

Let S^p be purely propositional as in Lemma 8. Using completeness of ILM with respect to Veltman frames, it is easy to see that

$$\text{ILM} \vdash \square\perp \rightarrow (\top \triangleright S(A, B) \leftrightarrow \top \triangleright S^p(A, B)). \tag{25}$$

By (25), and arithmetical soundness of ILM, we have

$$\text{PA} \vdash \square\perp \rightarrow (\top \triangleright S^*(A^*, B^*) \leftrightarrow \top \triangleright (S^p)^*(A^*, B^*)),$$

where $*$ is any arithmetical realisation for ILM. Combining with (24), we get for any arithmetical realization $*$,

$$\text{PA} \vdash \square\perp \rightarrow (\top \triangleright (S^p)^*(A^*, B^*) \leftrightarrow (\top \triangleright A^*) \wedge (\top \triangleright B^*)). \tag{26}$$

Since S^p is purely propositional, there are 16 possibilities. We will show that in each case, there are \mathcal{L} -formulas A and B such that

$$\text{PA} \vdash (\top \triangleright S^p(A, B) \leftrightarrow (\top \triangleright A) \wedge (\top \triangleright B)) \rightarrow \square\perp.$$

Combining the above with (26), we get $\text{PA} \vdash \square\perp \rightarrow \square\perp$ in all cases, a contradiction.

Argue in PA, assuming $\top \triangleright S^p(A, B) \leftrightarrow (\top \triangleright A) \wedge (\top \triangleright B)$. We use reasoning warranted by ILM to show $\top \triangleright \perp$, which is equivalent to $\square\perp$.

1. $S^P(A, B) = \top$. We have $\top \triangleright \top$, and therefore by assumption $(\rightarrow) \top \triangleright \perp$ (taking \perp for A).
2. $S^P(A, B) = \perp$. Take \top for A and B . We have $\top \triangleright \top$, and therefore by assumption $(\leftarrow) \top \triangleright \perp$.
3. $S^P(A, B) = A$. Take \top for A and \perp for B . We have $\top \triangleright A$, and thus by assumption $\top \triangleright B$, i.e. $\top \triangleright \perp$.
4. $S^P(A, B) = B$. Like the previous case, but take \top for B and \perp for A .
5. $S^P(A, B) = \neg A$. Take \perp for A . Then $\top \triangleright \neg A$, whence by assumption $\top \triangleright A$, i.e. $\top \triangleright \perp$.
6. $S^P(A, B) = \neg B$. Like the previous case, but take \perp for B .
7. $S^P(A, B) = A \wedge B$. Take as A an Orey sentence O , and for B the sentence $\neg O$. Then $\top \triangleright O$ and $\top \triangleright \neg O$, by the fact that O is an Orey sentence, so by assumption $\top \triangleright O \wedge \neg O$, i.e. $\top \triangleright \perp$.
8. $S^P(A, B) = A \vee B$. Take \perp for A and \top for B . Then $\top \triangleright A \vee B$, so by assumption $\top \triangleright A$, i.e. $\top \triangleright \perp$.
9. $S^P(A, B) = \neg A \wedge \neg B$. Take \perp for both A and B . Then $\top \triangleright \neg A \wedge \neg B$, so by assumption $\top \triangleright A$, i.e. $\top \triangleright \perp$.
10. $S^P(A, B) = \neg A \vee \neg B$. Take \perp for A . Then $\top \triangleright \neg A \vee \neg B$, so by assumption $\top \triangleright A$, i.e. $\top \triangleright \perp$.
11. $S^P(A, B) = A \wedge \neg B$. Take \top for A , and \perp for B . Then $\top \triangleright A \wedge \neg B$, so by assumption $\top \triangleright B$, i.e. $\top \triangleright \perp$.
12. $S^P(A, B) = \neg A \wedge B$. Like the previous case, but take \perp for A and \top for B .
13. $S^P(A, B) = A \rightarrow B$. Take \top for B and \perp for A . Then $\top \triangleright A \rightarrow B$, so by assumption $\top \triangleright A$, i.e. $\top \triangleright \perp$.
14. $S^P(A, B) = B \rightarrow A$. Like the previous case, but take \perp for B and \top for A .
15. $S^P(A, B) = A \leftrightarrow B$. Take \perp for both A and B . Then $\top \triangleright A \leftrightarrow B$, so also $\top \triangleright A$, i.e. $\top \triangleright \perp$ by assumption.
16. $S^P(A, B) = \neg(A \leftrightarrow B)$. Take \top for A and \perp for B . Then $\top \triangleright \neg(A \leftrightarrow B)$, so by assumption also $\top \triangleright B$, i.e. $\top \triangleright \perp$.

Remark 3 The above proof proceeds by showing that if suprema were definable in ILM, then $\Box\Box\perp \rightarrow \Box\perp$ would be provable in PA. Note that since $\Box\Box\perp \rightarrow \Box\perp$ is true, the same argument cannot be used to show that the existence of the supremum is not expressible in ILM^ω , the modal logic of interpretability statements that are true (in the standard model).

Question 4 Are interpretability suprema definable in ILM^ω ?

A.2 A non-commutative supremum

We construct a supremum implementation τ , such that $\tau(A, B) \leftrightarrow \tau(B, A)$ is not always provable in PA. Let σ be any supremum implementation. Let $\tau(A, B)$ be the formula

$$(A \leq B \rightarrow \sigma(A, B) \wedge \Box\neg\sigma(A, B)) \wedge (B < A \rightarrow \sigma(A, B) \vee \Diamond\sigma(A, B)),$$

where A and B are regarded as internal variables ranging over sentences. By $A \leq B$ we mean that the gödelnumber of A is smaller than or equal to the gödelnumber of B .

We show first that τ is a supremum implementation. Argue in PA . We first show $\tau(A, B) \triangleright A$. In case $A \leq B$, we have $\Box(\tau(A, B) \rightarrow \sigma(A, B))$, and thus also $\tau(A, B) \triangleright \sigma(A, B)$. Since $\sigma(A, B) \triangleright A$, also $\tau(A, B) \triangleright A$ by transitivity of \triangleright . In case $B < A$, we have $\Box(\tau(A, B) \rightarrow \sigma(A, B) \vee \Diamond\sigma(A, B))$, and so also $\tau(A, B) \triangleright \sigma(A, B) \vee \Diamond\sigma(A, B)$. Since by axiom **J5** of **ILM**, $\Diamond\sigma(A, B) \triangleright \sigma(A, B)$ we get

$$\sigma(A, B) \vee \Diamond\sigma(A, B) \triangleright \sigma(A, B)$$

by axiom **J3** of **ILM**, and so $\sigma(A, B) \vee \Diamond\sigma(A, B) \triangleright A$ by the fact that $\sigma(A, B) \triangleright A$, and transitivity of \triangleright . Thus $\tau(A, B) \triangleright A$ follows again by transitivity of \triangleright . The argument for $\tau(A, B) \triangleright B$ is similar. It remains to show that if C is such that $C \triangleright A$ and $C \triangleright B$, then $C \triangleright \tau(A, B)$. We clearly have $C \triangleright \sigma(A, B)$ for any such C . Consider first the case that $A \leq B$. Then

$$\Box(\tau(A, B) \leftrightarrow \sigma(A, B) \wedge \Box\neg\sigma(A, B)).$$

Since $D \triangleright D \wedge \Box\neg D$ is a theorem of **ILM**, we have that $\sigma(A, B) \triangleright \sigma(A, B) \wedge \Box\neg\sigma(A, B)$, and therefore also $\sigma(A, B) \triangleright \tau(A, B)$ by transitivity of \triangleright . In case $B < A$, we have that $\Box(\tau(A, B) \leftrightarrow \sigma(A, B) \vee \Diamond\sigma(A, B))$, and so $\Box(\sigma(A, B) \rightarrow \tau(A, B))$, whence again $\sigma(A, B) \triangleright \tau(A, B)$. Thus if $C \triangleright \sigma(A, B)$, then also $C \triangleright \tau(A, B)$.

Finally, we show τ is not commutative with respect to provability. For that, let A and B be distinct \mathcal{L} -sentences (thus under a reasonable gödelnumbering, their gödelnumbers are also different). We assume without loss of generality that $A < B$, and suppose for a contradiction that $\text{PA} \vdash \tau(A, B) \leftrightarrow \tau(B, A)$. By definition of τ , this means that

$$\text{PA} \vdash \sigma(A, B) \wedge \Box\neg\sigma(A, B) \leftrightarrow \sigma(A, B) \vee \Diamond\sigma(A, B). \tag{27}$$

It follows from (27) by propositional reasoning that $\text{PA} \vdash \Box\neg\sigma(A, B)$, and thus $\text{PA} \vdash \neg\sigma(A, B)$. But this means that $\sigma(A, B)$, i.e. the supremum of A and B , is in the degree of $[\perp]$, contradicting our assumption.

B Varieties of supremum implementations

B.1 Obtaining supremum implementations

We show how to turn Švejdar’s construction into a supremum implementation as in Definition 3. The case of Visser’s construction is similar. Let $\text{Sub}(x, y, z)$ be an intensionally correct formula representing in $\text{I}\Delta_0+\text{exp}$ substitution of numerals for free variables in formulas. Thus for all m, n , and all \mathcal{L} -formulas A ,

$$\text{I}\Delta_0+\text{exp} \vdash \text{Sub}(\ulcorner A(u, v) \urcorner, \bar{m}, \bar{n}) = \ulcorner A(\bar{m}, \bar{n}) \urcorner,$$

By Theorem 1, there is a formula σ with two free variables, and such that

$$I\Delta_0 + \text{exp} \vdash \sigma(A, B) \leftrightarrow \forall x (\diamond_x \text{Sub}(\ulcorner \sigma(u, v) \urcorner, A, B) \rightarrow \diamond_x A \wedge \diamond_x B),$$

where A and B are seen as internal variables ranging over \mathcal{L} -sentences. By properties of Sub,

$$I\Delta_0 + \text{exp} \vdash \sigma(A, B) \leftrightarrow \forall x (\diamond_x \sigma(A, B) \rightarrow \diamond_x A \wedge \diamond_x B).$$

Thus the formula σ is exactly what we were looking for.

B.2 Unique explicit fixed points

We prove Theorems 7 and 8, repeated here as Theorems 14 and 15. We start with the following lemma.

Lemma 9 *Let $\varepsilon(A, B)$ be the formula*

$$(\Box_0 \neg A \vee \Box_0 \neg B \rightarrow \Box_0 \perp) \wedge \forall x (\Box_{x+1} \neg A \vee \Box_{x+1} \neg B \rightarrow \Box_{x+1} (\Box_x \neg A \vee \Box_x \neg B)),$$

where A and B are regarded as internal variables ranging over \mathcal{L} -sentences. Then

- i. $I\Delta_0 + \text{exp} \vdash \Box_0 \neg \varepsilon(A, B) \leftrightarrow \Box_0 \perp$
- ii. $I\Delta_0 + \text{exp} \vdash \forall x (\Box_{x+1} \neg \varepsilon(A, B) \leftrightarrow \Box_{x+1} (\Box_x \neg A \vee \Box_x \neg B))$

Proof For the nontrivial direction of i, reason in $I\Delta_0 + \text{exp}$, assuming $\Box_0 \neg \varepsilon(A, B)$, i.e.

$$\Box_0 (((\Box_0 \neg A \vee \Box_0 \neg B) \wedge \diamond_0 \top) \vee \exists x ((\Box_{x+1} \neg A \vee \Box_{x+1} \neg B) \wedge \neg \Box_{x+1} (\Box_x \neg A \vee \Box_x \neg B))). \tag{28}$$

(28) implies $\Box_0 (\diamond_0 \top \vee \exists x \neg \Box_{x+1} (\Box_x \neg A \vee \Box_x \neg B))$. Since for any C , $\neg \Box_{x+1} C$ implies $\neg \Box_0 \perp$ by monotonicity, it follows from (28) that $\Box_0 \diamond_0 \top$, and therefore $\Box_0 \perp$ by Löb’s principle.

For ii, reason in $I\Delta_0 + \text{exp}$, and let x be arbitrary. Assume first

$$\Box_{x+1} (\Box_x \neg A \vee \Box_x \neg B). \tag{29}$$

To show that $\Box_{x+1} \neg \varepsilon(A, B)$, reason in \Box_{x+1} . Suppose for a contradiction that $\varepsilon(A, B)$. By (29), let $\mu \leq x$ be the minimal number s.t. $\Box_\mu \neg A \vee \Box_\mu \neg B$. If $\mu = 0$, then $\Box_0 \perp$ follows by $\varepsilon(A, B)$, which is a contradiction since we have $\neg \Box_0 \perp$ by reflection. Thus $\mu = y + 1$ for some y , whence $\varepsilon(A, B)$ now gives us

$$\Box_{y+1} (\Box_y \neg A \vee \Box_y \neg B). \tag{30}$$

Since $y + 1 < x + 1$, from (30) we get $\Box_y \neg A \vee \Box_y \neg B$ by reflection, contradicting the minimality of $y + 1$. We conclude that $\neg \varepsilon(A, B)$.

For the other direction, assume $\Box_{x+1}\neg\varepsilon(A, B)$. Reason in \Box_{x+1} . Note that $\neg\varepsilon(A, B)$ is the sentence

$$((\Box_0\neg A \vee \Box_0\neg B) \wedge \Diamond_0\top) \vee \exists y ((\Box_{y+1}\neg A \vee \Box_{y+1}\neg B) \wedge \neg\Box_{y+1}(\Box_y\neg A \vee \Box_y\neg B)). \tag{31}$$

If the first disjunct holds or some $y < x$ witnesses the second disjunct, then by monotonicity we have that $\Box_x\neg A \vee \Box_x\neg B$. If the smallest witness of the second disjunct is some $y \geq x$, then (by monotonicity) $\neg\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)$. Thus $(\Box_x\neg A \vee \Box_x\neg B) \vee \neg\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)$, or in other words

$$\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B) \rightarrow \Box_x\neg A \vee \Box_x\neg B.$$

Leaving the inner \Box_{x+1} -world, we have shown: $\Box_{x+1}(\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B) \rightarrow \Box_x\neg A \vee \Box_x\neg B)$. By Löb’s Theorem we can conclude $\Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)$ as required. \square

Theorem 14 *Let $\varepsilon(A, B)$ be the formula*

$$(\Box_0\neg A \vee \Box_0\neg B \rightarrow \Box_0\perp) \wedge \forall x (\Box_{x+1}\neg A \vee \Box_{x+1}\neg B \rightarrow \Box_{x+1}(\Box_x\neg A \vee \Box_x\neg B)),$$

where A and B are regarded as internal variables ranging over \mathcal{L} -sentences. Then

- i. $\text{I}\Delta_0+\text{exp} \vdash \varepsilon(A, B) \leftrightarrow \forall x (\Diamond_x \varepsilon(A, B) \rightarrow \Diamond_x A \wedge \Diamond_x B)$
- ii. For any ϑ , if $\text{PA} \vdash \vartheta \leftrightarrow \forall x (\Diamond_x \vartheta \rightarrow \Diamond_x A \wedge \Diamond_x B)$, then $\text{PA} \vdash \vartheta \leftrightarrow \varepsilon(A, B)$

Proof For i, argue in $\text{I}\Delta_0+\text{exp}$, and assume $\varepsilon(A, B)$. Using contraposition, it suffices to show that for all x , $\Box_x\neg A \vee \Box_x\neg B$ implies $\Box_x\neg\varepsilon(A, B)$. Suppose $\Box_x\neg A \vee \Box_x\neg B$. If $x = 0$, then $\varepsilon(A, B)$ implies $\Box_0\perp$, and hence trivially $\Box_0\neg\varepsilon(A, B)$. If $x = y + 1$ for some y , then by $\varepsilon(A, B)$, we have $\Box_{y+1}(\Box_y\neg A \vee \Box_y\neg B)$ and hence $\Box_{y+1}\neg\varepsilon(A, B)$, i.e. $\Box_x\neg\varepsilon(A, B)$ by Lemma 9. For the other direction assume $\neg\varepsilon(A, B)$, i.e. the sentence in (31). By Lemma 9 i, the first disjunct implies $(\Box_0\neg A \wedge \Box_0\neg B) \wedge \Diamond_0\varepsilon(A, B)$, and by Lemma 9 ii the second disjunct implies $\exists x ((\Box_{x+1}\neg A \vee \Box_{x+1}\neg B) \wedge \Diamond_{x+1}\varepsilon(A, B))$. In either case $\exists x (\Box_x\neg A \vee \Box_x\neg B \wedge \Diamond_x\varepsilon(A, B))$, i.e. $\neg\forall x (\Diamond_x\varepsilon(A, B) \rightarrow \Diamond_x A \wedge \Diamond_x B)$, which is what we wanted to show.

For ii, it now suffices to show that any two fixed points of the formula $\forall x (\Diamond_x Y \rightarrow (\Diamond_x A \wedge \Diamond_x B))$ are provably equivalent in PA . Let n be s.t.

$$\begin{aligned} \text{I}\Sigma_n \vdash \vartheta &\leftrightarrow \forall x (\Box_x\neg A \vee \Box_x\neg B \rightarrow \Box_x\neg\vartheta) \quad \text{and} \\ \text{I}\Sigma_n \vdash \sigma &\leftrightarrow \forall x (\Box_x\neg A \vee \Box_x\neg B \rightarrow \Box_x\neg\sigma). \end{aligned}$$

We show that $\text{I}\Sigma_n \vdash \sigma \leftrightarrow \vartheta$. Using Löb’s Theorem, it suffices to show $\text{I}\Sigma_n \vdash \Box_n(\vartheta \leftrightarrow \sigma) \rightarrow (\sigma \leftrightarrow \vartheta)$. We argue in $\text{I}\Sigma_n$ and assume $\Box_n(\vartheta \leftrightarrow \sigma)$. By monotonicity and validity of \mathbf{K} for \Box_n , we have for all $x \geq n$:

$$\Box_x\neg\vartheta \leftrightarrow \Box_x\neg\sigma. \tag{32}$$

Assume σ . In order to show ϑ , it suffices to show $\forall x (\Box_x \neg A \vee \Box_x \neg B \rightarrow \Box_x \neg \vartheta)$. Assume $\Box_x \neg A \vee \Box_x \neg B$. By σ , we get $\Box_x \neg \sigma$. By essential reflexivity, it must be that $x \geq n$, and thus $\Box_x \neg \vartheta$ follows by (32). \square

We now turn our attention to Visser’s supremum. We first prove a few lemmas.

Lemma 10 *Suppose that $I\Sigma_n \vdash \vartheta \leftrightarrow \forall x (\Box_x \vartheta \rightarrow \Box_x \neg A \vee \Box_x \neg B)$. Then*

$$I\Sigma_n \vdash \forall x \geq n \Box_x (\forall y < x (\Box_y \vartheta \rightarrow \Box_y \neg A \vee \Box_y \neg B)).$$

Proof By necessitation,

$$I\Sigma_n \vdash \Box_n (\vartheta \leftrightarrow \forall x (\Box_x \vartheta \rightarrow \Box_x \neg A \vee \Box_x \neg B)),$$

whence we can use ϑ and $\forall x (\Box_x \vartheta \rightarrow \Box_x \neg A \vee \Box_x \neg B)$ interchangeably in \Box_x if $x \geq n$.

$$\begin{aligned} I\Sigma_n \vdash x \geq n \rightarrow \Box_x \forall y < x (\Box_y \vartheta \rightarrow \vartheta) & \quad \text{(reflection)} \\ \rightarrow \Box_x \forall y < x (\Box_y \vartheta \rightarrow \forall z (\Box_z \vartheta \rightarrow \Box_z \neg A \vee \Box_z \neg B)) & \quad \text{(fixed point version of } \vartheta) \\ \rightarrow \Box_x \forall y < x (\Box_y \vartheta \rightarrow \Box_y \neg A \vee \Box_y \neg B) & \end{aligned}$$

\square

Lemma 11 *Suppose that $I\Sigma_n \vdash \vartheta \leftrightarrow \forall x (\Box_x \vartheta \rightarrow \Box_x \neg A \vee \Box_x \neg B)$. Then*

$$I\Sigma_n \vdash \forall x \geq n (\Box_x \vartheta \leftrightarrow \Box_x (\Box_x \neg A \vee \Box_x \neg B)).$$

Proof

$$\begin{aligned} I\Sigma_n \vdash x \geq n \rightarrow (\Box_x \vartheta \rightarrow \Box_x \forall y (\Box_y \vartheta \rightarrow \Box_y \neg A \vee \Box_y \neg B)) & \quad \text{(fixed point version of } \vartheta) \\ \rightarrow \Box_x (\Box_x \vartheta \rightarrow \Box_x \neg A \vee \Box_x \neg B) & \quad \text{(instantiating } \forall) \\ \rightarrow (\Box_x \Box_x \vartheta \rightarrow \Box_x (\Box_x \neg A \vee \Box_x \neg B)) & \quad \text{(K-axiom)} \\ \rightarrow \Box_x (\Box_x \neg A \vee \Box_x \neg B) & \quad \text{(since } \Box_x \vartheta \rightarrow \Box_x \Box_x \vartheta) \\ \rightarrow \Box_x \forall y \geq x (\Box_y \neg A \vee \Box_y \neg B) & \quad \text{(monotonicity)} \\ \rightarrow \Box_x \forall y \geq x (\Box_y \vartheta \rightarrow \Box_y \neg A \vee \Box_y \neg B) & \\ \rightarrow \Box_x \forall y (\Box_y \vartheta \rightarrow \Box_y \neg A \vee \Box_y \neg B) & \quad \text{(Lemma 10)} \\ \rightarrow \Box_x \vartheta & \quad \text{(fixed point version of } \vartheta) \end{aligned}$$

Lemma 12 *Let $\varepsilon(A, B)$ be the formula $\exists x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \wedge \Diamond_x A \wedge \Diamond_x B)$, where A and B are regarded as internal variables ranging over sentences. Then*

$$I\Delta_0 + \text{exp} \vdash \forall x (\Box_x \neg \varepsilon(A, B) \leftrightarrow \Box_x (\Box_x \neg A \vee \Box_x \neg B)).$$

Proof

$$I\Delta_0 + \text{exp} \vdash \Box_x \neg \varepsilon(A, B) \rightarrow \Box_x \forall y (\Box_y (\Box_y \neg A \vee \Box_y \neg B) \rightarrow \Box_y \neg A \vee \Box_y \neg B) \quad (33)$$

$$\rightarrow \Box_x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \rightarrow \Box_x \neg A \vee \Box_x \neg B) \quad (34)$$

$$\rightarrow \Box_x (\Box_x \neg A \vee \Box_x \neg B) \quad (35)$$

$$\rightarrow \Box_x \forall y \geq x (\Box_y \neg A \vee \Box_y \neg B) \tag{36}$$

$$\rightarrow \Box_x \forall y \geq x (\Box_y (\Box_y \neg A \vee \Box_y \neg B) \rightarrow \Box_y \neg A \vee \Box_y \neg B) \wedge \tag{37}$$

$$\Box_x \forall y < x (\Box_y (\Box_y \neg A \vee \Box_y \neg B) \rightarrow \Box_y \neg A \vee \Box_y \neg B) \tag{38}$$

$$\rightarrow \Box_x \forall y (\Box_y (\Box_y \neg A \vee \Box_y \neg B) \rightarrow \Box_y \neg A \vee \Box_y \neg B) \tag{39}$$

$$\rightarrow \Box_x \neg \varepsilon(A, B) \tag{40}$$

Step (35) is Löb’s Theorem, step (36) monotonicity, step (37) the previous step together with propositional logic, and step (38) reflection. \square

Theorem 15 *Let $\varepsilon(A, B)$ be the formula*

$$\exists x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B)). \tag{41}$$

where A and B are regarded as internal variables ranging over \mathcal{L} -sentences. Then

- i. $\text{PA} \vdash \varepsilon(A, B) \leftrightarrow \exists x (\Box_x \neg \varepsilon(A, B) \wedge (\Diamond_x A \wedge \Diamond_x B))$
- ii. For any ϑ , if $\text{PA} \vdash \vartheta \leftrightarrow \exists x (\Box_x \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$, then $\text{PA} \vdash \vartheta \leftrightarrow \varepsilon(A, B)$

Proof Item i is an immediate consequence of Lemma 12, together with the definition of $\varepsilon(A, B)$. For ii, assume $\text{I}\Sigma_n \vdash \vartheta \leftrightarrow \exists x (\Box_x \neg \vartheta \wedge (\Diamond_x A \wedge \Diamond_x B))$. By contraposition,

$$\text{I}\Sigma_n \vdash \neg \vartheta \leftrightarrow \forall x (\Box_x \neg \vartheta \rightarrow (\Box_x \neg A \vee \Box_x \neg B)).$$

By Lemma 11 and reflection,

$$\text{I}\Sigma_n \vdash \neg \vartheta \leftrightarrow \forall x \geq n (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \rightarrow \Box_x \neg A \vee \Box_x \neg B). \tag{42}$$

On the other hand, we have by reflection

$$\text{I}\Sigma_n \vdash \forall x < n (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \rightarrow \Box_x \neg A \vee \Box_x \neg B). \tag{43}$$

Combining (42) and (43), we have $\text{I}\Sigma_n \vdash \neg \vartheta \leftrightarrow \forall x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \rightarrow \Box_x \neg A \vee \Box_x \neg B)$, whence by contraposition, $\text{I}\Sigma_n \vdash \vartheta \leftrightarrow \exists x (\Box_x (\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B))$, i.e. $\text{I}\Sigma_n \vdash \vartheta \leftrightarrow \varepsilon(A, B)$ as required. \square

B.3 Extensionality

We prove Theorem 9 from Sect. 4. We start with a lemma.

Lemma 13 ($\text{I}\Delta_0 + \text{exp}$) *Let $\varepsilon(A, B)$ be as in Theorem 14. Then*

$$\text{I}\Sigma_{n+1} \vdash \varepsilon(A, B) \leftrightarrow \tag{44}$$

$$(\Diamond_n A \wedge \Diamond_n B) \wedge \forall x \geq n (\Diamond_{x+1} (\Diamond_x A \wedge \Diamond_x B) \rightarrow \Diamond_{x+1} A \wedge \Diamond_{x+1} B) \tag{45}$$

Proof By propositional reasoning, we see that $\varepsilon(A, B)$ is equivalent to

$$(\diamond_0 \top \rightarrow \diamond_0 A \wedge \diamond_0 B) \wedge \forall x (\diamond_{x+1}(\diamond_x A \wedge \diamond_x B) \rightarrow \diamond_{x+1} A \wedge \diamond_{x+1} B). \quad (46)$$

Argue in $I\Sigma_{n+1}$, assuming (46). By reflection, $\diamond_0 \top$, and therefore $\diamond_0 A \wedge \diamond_0 B$ from the first conjunct of (46). Now let $x < n$. By reflection for $I\Sigma_{x+1}$ (remember that we are reasoning inside $I\Sigma_{n+1}$), we have that

$$\Box_{x+1}(\Box_x \neg A \vee \Box_x \neg B) \rightarrow (\Box_x \neg A \vee \Box_x \neg B),$$

i.e. by contraposition,

$$\diamond_x A \wedge \diamond_x B \rightarrow \diamond_{x+1}(\diamond_x A \wedge \diamond_x B). \quad (47)$$

Since we have $\diamond_0 A \wedge \diamond_0 B$, we can apply (47) and the second conjunct of (46) to get $\diamond_{x+1} A \wedge \diamond_{x+1} B$ for all $x < n$, and thus $\diamond_n A \wedge \diamond_n B$. The other conjunct of (45) clearly follows from (46). This finishes the proof from (44) to (45). For the other direction, we note that $\diamond_n A \wedge \diamond_n B$ implies $\diamond_0 A \wedge \diamond_0 B$, taking care of the first conjunct of (46). For the other conjunct, it suffices to show that for $1 \leq x < n$,

$$\diamond_{x+1}(\diamond_x A \wedge \diamond_x B) \rightarrow \diamond_{x+1} A \wedge \diamond_{x+1} B. \quad (48)$$

This follows because we have $\diamond_{x+1} A \wedge \diamond_{x+1} B$ for all such x by assumption. \square

Theorem 16 *Let $\varepsilon(A, B)$ be as in Theorem 14. Then*

$$I\Delta_0 + \text{exp} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \rightarrow \Box(\varepsilon(A, B) \leftrightarrow \varepsilon(A', B'))$$

Proof Clearly it suffices to show that

$$I\Delta_0 + \text{exp} \vdash \Box_x(A \leftrightarrow A') \wedge \Box_x(B \leftrightarrow B') \rightarrow \Box_{x+1}(\varepsilon(A, B) \leftrightarrow \varepsilon(A', B')).$$

Argue in $I\Delta_0 + \text{exp}$, assuming

$$\Box_n(A \leftrightarrow A') \wedge \Box_n(B \leftrightarrow B'). \quad (49)$$

Now argue inside $I\Sigma_{n+1}$. Using Lemma 13 it suffices to show that

$$(\diamond_n A \wedge \diamond_n B) \wedge \forall x \geq n (\diamond_{x+1}(\diamond_x A \wedge \diamond_x B) \rightarrow \diamond_{x+1} A \wedge \diamond_{x+1} B) \quad (50)$$

if and only if

$$(\diamond_n A' \wedge \diamond_n B') \wedge \forall x \geq n (\diamond_{x+1}(\diamond_x A' \wedge \diamond_x B') \rightarrow \diamond_{x+1} A' \wedge \diamond_{x+1} B'). \quad (51)$$

Since (49) is a Σ_1 -sentence, its truth carries over to our current surroundings inside $I\Sigma_{n+1}$. By monotonicity and modal reasoning, it follows that for all $x \geq n$, $\Box_x A \leftrightarrow$

$\Box_x A'$ and $\Diamond_x A \leftrightarrow \Diamond_x A'$; similarly for B and B' . Given that, the equivalence of (50) and (51) is clear. □

Theorem 17 *Let $\varepsilon(A, B)$ be as in Theorem 15. Then*

$$I\Delta_0 + \mathbf{exp} \vdash \Box(A \leftrightarrow A') \wedge \Box(B \leftrightarrow B') \rightarrow \Box(\varepsilon(A, B) \leftrightarrow \varepsilon(A', B'))$$

Proof We shall show that

$$I\Delta_0 + \mathbf{exp} \vdash \Box_x(A \leftrightarrow A') \wedge \Box_x(B \leftrightarrow B') \rightarrow \Box_x(\varepsilon(A, B) \leftrightarrow \varepsilon(A', B')).$$

Argue in $I\Delta_0 + \mathbf{exp}$, assuming

$$\Box_n(A \leftrightarrow A') \wedge \Box_n(B \leftrightarrow B') \tag{52}$$

Argue in $I\Sigma_n$. We want to show that

$$\exists x (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B)). \tag{53}$$

if and only if

$$\exists x (\Box_x(\Box_x \neg A' \vee \Box_x \neg B') \wedge (\Diamond_x A' \wedge \Diamond_x B')). \tag{54}$$

Since (52) is a Σ_1 -sentence, its truth carries over to the world inside $I\Sigma_n$. By monotonicity and modal reasoning, it follows that for all $x \geq n$, $\Box_x A \leftrightarrow \Box_x A'$ and $\Diamond_x A \leftrightarrow \Diamond_x A'$, and also $\Box_x(\Box_x A \leftrightarrow \Box_x A')$ (again using that $\Box_x A \leftrightarrow \Box_x A'$ is Σ_1). Similarly for B and B' . Given that, the equivalence of (53) and (54) is clear. □

B.4 Failure of monotonicity for Visser’s supremum

We prove Theorem 10 from Sect. 4.

Theorem 18 *Let $\varepsilon(A, B)$ be $\exists x (\Box_x(\Box_x \neg A \vee \Box_x \neg B) \wedge (\Diamond_x A \wedge \Diamond_x B))$, where A and B are seen as internal variables over \mathcal{L} -sentences. The formula ε is not monotone.*

Proof It suffices to show that there is a sentence A for which

$$PA \not\vdash \varepsilon(A, A) \rightarrow \varepsilon(\top, \top). \tag{55}$$

In fact we can take for A the sentence $\varepsilon(\top, \top)$. We note that:

$$\varepsilon(\top, \top) = \exists x (\Box_x \Box_x \perp \wedge \Diamond_x \top) \tag{56}$$

$$\varepsilon(\varepsilon(\top, \top), \varepsilon(\top, \top)) = \exists x (\Box_x \Box_x \neg \varepsilon(\top, \top) \wedge \Diamond_x \varepsilon(\top, \top)) \tag{57}$$

Recall that by Lemma 12,

$$I\Delta_0 + \mathbf{exp} \vdash \forall x (\Box_x \neg \varepsilon(\top, \top) \leftrightarrow \Box_x \Box_x \perp) \tag{58}$$

Using (58), we see that, verifiably in $I\Delta_0+\text{exp}$, (57) is equivalent to

$$\exists x (\Box_x \Box_x \Box_x \perp \wedge \Diamond_x \Diamond_x \top) \quad (59)$$

Thus in order to show (55) it suffices to show

$$\text{PA} \not\vdash \exists x (\Box_x \Box_x \Box_x \perp \wedge \Diamond_x \Diamond_x \top) \rightarrow \exists x (\Box_x \Box_x \perp \wedge \Diamond_x \top).$$

Suppose for a contradiction that $\exists x (\Box_x \Box_x \Box_x \perp \wedge \Diamond_x \Diamond_x \top) \rightarrow \exists x (\Box_x \Box_x \perp \wedge \Diamond_x \top)$ is provable in some $I\Sigma_n$. Argue in $I\Sigma_n$. From our assumption, it follows that in particular

$$(\Box_n \Box_n \Box_n \perp \wedge \Diamond_n \Diamond_n \top) \rightarrow \exists x (\Box_x \Box_x \perp \wedge \Diamond_x \top). \quad (60)$$

Assume $\Box_n \Box_n \Box_n \perp \wedge \Diamond_n \Diamond_n \top$, and let x be such that $\Box_x \Box_x \perp \wedge \Diamond_x \top$. Since we have reflection for all $m < n$, it must be that $x \geq n$. But if $x > n$, then $\Box_n \Box_n \Box_n \perp$ implies $\Box_x \perp$ by monotonicity and reflection. Thus it must be that $x = n$. We now exit the world inside $I\Sigma_n$. We have shown:

$$I\Sigma_n \vdash (\Box_n \Box_n \Box_n \perp \wedge \Diamond_n \Diamond_n \top) \rightarrow (\Box_n \Box_n \perp \wedge \Diamond_n \top);$$

It follows by propositional logic that

$$I\Sigma_n \vdash \Box_n \Box_n \Box_n \perp \rightarrow \Box_n \Box_n \perp.$$

By Löb's Theorem, we now have $I\Sigma_n \vdash \Box_n \Box_n \perp$, a contradiction. \square

References

- Berarducci, A.: The interpretability logic of Peano arithmetic. *J. Symb. Logic* **55**, 1059–1089 (1990)
- Boolos, G.: *The Logic of Provability*. Cambridge University Press, Cambridge (1993)
- Ehrenfeucht, A., Feferman, S.: Representability of recursively enumerable sets in formal theories. *Archiv für mathematische Logik und Grundlagenforschung* **5**(1), 37–41 (1961). doi:[10.1007/BF01977641](https://doi.org/10.1007/BF01977641)
- Feferman, S.: Arithmetization of metamathematics in a general setting. *Fundam. Math.* **49**, 35–92 (1960)
- Guaspari, D.: Partially conservative extensions of arithmetic. *Trans. Am. Math. Soc.* **254**, 47–68 (1979)
- Hájek, P.: On interpretability in set theories I. *Commun. Math. Univ. Carolinae* **12**, 73–79 (1971)
- Hájek, P., Pudlák, P.: *Metamathematics of First-Order Arithmetic*. Springer, Berlin (1998)
- Hájková, M., Hájek, P.: On interpretability in theories containing arithmetic. *Fundam. Math.* **76**(2), 131–137 (1972). <http://eudml.org/doc/214455>
- Henk, P., Pakhomov, F.: Slow and Ordinary Provability for Peano Arithmetic. ArXiv e-prints (2016)
- Henk, P., Shavrukov, V. Yu.: A Solovay function for the least 1-inconsistent subtheory of PA. Prepublication Series. Report No. PP-2016-31. ILLC, University of Amsterdam (2016)
- Hilbert, D., Bernays, P.: *Grundlagen der Mathematik II*. Springer, Berlin (1939). Second edition: 1970
- Jeroslow, R.G.: Consistency statements in formal theories. *Fundam. Math.* **72**, 17–40 (1971)
- de Jongh, D., Jumelet, M., Montagna, F.: On the proof of Solovay's theorem. *Stud. Log.* **50**, 51–70 (1991)
- de Jongh, D., Veltman, F.: Provability logics for relative interpretability. In: *Mathematical logic, Proceedings of the Heyting 1988 summer school in Varna, Bulgaria*, pp. 31–42. Plenum Press, Boston (1990)

15. Lindström, P.: On certain lattices of degrees of interpretability. *Notre Dame J. Formal Logic* **25**(2), 127–140 (1984). doi:[10.1305/ndjfl/1093870573](https://doi.org/10.1305/ndjfl/1093870573)
16. Lindström, P.: Some results on interpretability. In: F. Jensen, M. B.H., M. K.K. (eds.) *Proceedings of the 5th Scandinavian Logic Symposium 1979*, pp. 329–361. Aalborg University Press, Aalborg (1979)
17. Lindström, P.: On Parikh provability—an exercise in modal logic. In: H. Lagerlund, S. Lindström, R. Sliwinski (eds.) *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, Uppsala Philosophical Studies 53, pp. 279–288 (2006)
18. Löb, M.: Solution of a problem of Leon Henkin. *J. Symb. Logic* **20**, 115–118 (1955)
19. Montagna, F.: On the algebraization of a Feferman’s predicate. *Stud. Log.* **37**, 221–236 (1978)
20. Montague, R.: Theories incomparable with respect to relative interpretability. *J. Symb. Logic* **27**(2), 195–211 (1962). <http://projecteuclid.org/euclid.jsl/1183734432>
21. Ono, H.: Reflection principles in fragments of Peano arithmetic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **33**, 317–333 (1987)
22. Orey, S.: Relative interpretations. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **7**, 146–153 (1961)
23. Parikh, R.: Existence and feasibility in arithmetic. *J. Symb. Logic* **36**(3), 494–508 (1971). <http://www.jstor.org/stable/2269958>
24. Shavrukov, V.: The logic of relative interpretability over Peano arithmetic (in Russian). Tech. Rep. No.5, Steklov Mathematical Institute, Moscow (1988)
25. Shavrukov, V.: A smart child of Peano’s. *Notre Dame J. Formal Logic* **35**, 161–185 (1994)
26. Smoryński, C.: Arithmetic analogues of McAloon’s unique Rosser sentences. *Arch. Math. Logic* **28**, 1–21 (1989)
27. Solovay, R.: Provability interpretations of modal logic. *Israel J. Math.* **25**, 287–304 (1976)
28. Švejdar, V.: Degrees of interpretability. *Comment. Math. Univ. Carolinae* **19**, 789–813 (1978)
29. Tarski, A., Mostowski, A., Robinson, R.: *Undecidable Theories*. North-Holland, Amsterdam (1953)
30. Visser, A.: Peano’s smart children: A provability logical study of systems with built-in consistency. *Notre Dame J. Formal Logic* **30**, 161–196 (1989)
31. Visser, A.: An inside view of EXP. *J. Symb. Logic* **57**, 131–165 (1992)
32. Visser, A.: An Overview of Interpretability Logic. In: M. Kracht, M. de Rijke, H. Wansing, M. Zakharyashev (eds.) *Advances in Modal Logic*, vol 1, CSLI Lecture Notes, no. 87, pp. 307–359. Center for the Study of Language and Information, Stanford (1998)